心理学报 2022, Vol. 54, No. 5, 481−496
Acta Psychologica Sinica

https://doi.org/10.3724/SP.J.1041.2022.00481

反馈对自我欺骗的影响：来自 ERP 的证据*

范 伟[1,2] 任梦梦[1,2] 张文洁[2,3] 钟毅平[1,2]

([1]湖南师范大学教育科学学院心理系; [2]认知与人类行为湖南省重点实验室, 长沙 410081)
([3]湖南师范大学教育科学学院学前教育系, 长沙 410081)

摘 要 本研究主要探讨反馈类型和反馈效价对自我欺骗的影响。实验 1 采用前瞻范式探究不同效价的模糊反馈对自我欺骗的影响，结果发现，相比于无答案提示组，有答案提示组的被试在积极模糊反馈的条件下显著地提高测试 2 第二阶段的预测分数(自我欺骗的程度加重)。实验 2 探究不同效价的精确反馈对自我欺骗的影响，结果发现，相对于无答案提示组，有答案提示组的被试在消极精确反馈条件下显著地降低测试 2 第二阶段的预测分数。实验 3 采用 ERP 技术探究反馈类型和反馈效价影响自我欺骗的内在机制，结果发现，模糊反馈条件下的自我欺骗诱发较大的 P2 成分、诱发较小的 N1 和 N400 成分；以及发现自我欺骗在额区的效应较大。总之，这些研究结果表明可能积极反馈促进个体的自我欺骗产生，消极反馈削弱个体自我欺骗的产生；积极模糊反馈能够进一步促进了个体自我欺骗的产生。这些研究结果还表明自我欺骗产生的背后机制可能是自我膨胀。

关键词 自我欺骗，反馈类型，反馈效价，前瞻范式，事件相关电位
分类号 B849: C91; B845

1 问题提出

在日常生活中个体为了达到某种目的会自己欺骗自己，并且为了保护自己免受充满敌意或恶意世界的影响会说服自己去做一些与事实相反的事情(Taylor & Brown, 1988; Mele, 1997b)。有学者将自己欺骗自己称为“自我欺骗”，这是一种强化和保护自我不受威胁信息影响的内在过程(Smith et al., 2017)。自我欺骗是人类中普遍存在的一种个体对自我能力产生高估的错误信念，该信念会使个体产生积极的错觉并否认自己的不足，并且在面对明确证据时仍然保持的一种对自我的积极信念(Bandura, 2011; Gao, 2021; Mele, 1997a; 谭成慧 等, 2020)。有研究发现自我欺骗可以提高个体的主观幸福感和促进个体的心理健康，使个体持有更积极的心态(Sheridan et al., 2015; Smith et al., 2017; Surbey, 2011)。另外自我欺骗可以促进人际交往和合作，获得更多的社会效益等(Chance et al., 2011; Sloman et al., 2010; Surbey, 2011; Smith et al., 2017)。

随着学者们不断增加对自我欺骗的研究兴趣，如何测量自我欺骗的发生发展显得十分重要。自我欺骗提升量表(Self-Deception Enhancement test, SDE)通过分数的变化表示产生自我欺骗的程度(Egerton et al., 2010; Gur & Sackeim, 1979; Uziel & Cohen, 2020; 钟罗金, 莫雷, 2019)。声音识别范式(voice recognition paradigm)、疼痛范式(pain paradigm)以及点跟踪范式(dot-tracking task)，这些研究范式大多通过计算个体的主观报告或完成某种实验任务来反应自我欺骗的差异(Chance et al., 2011; Gur & Sackeim, 1979; Quattrone & Tversky, 1984; Sloman et al., 2010; 范伟 等, 2017)。但这些范式可能会在测量自我欺骗的过程中容易受到个体主观报告等额外变量的影响而造成结果的误差。而前瞻范式则不同，不但可以避免这些缺点，反而具有更多的优点。在该范式的测试 1 阶段，个体在有答案的提示下获得较高的分数，此类个体往往会提高在测试 2

收稿日期: 2020-09-09
* 国家社会科学基金后期资助项目(21FSHB002)。
范伟和任梦梦为本文共同第一作者，对本文贡献等同。
通信作者：钟毅平，E-mail: ypzhong@hunnu.edu.cn

阶段实验任务的预测得分；若个体在测试 2 的实际得分与预测得分产生差值，这代表自我欺骗的产生(Ren et al., 2018)。该范式的优点在于能够检验个体的自我欺骗如何影响下一阶段对实验任务的预期(Liu et al., 2019; Ren et al., 2018)。

有研究发现外在信息的传递和反馈影响个体自我欺骗的产生(Quattrone et al., 1984; Sharma et al., 2010; Sloman et al., 2010)。个体在面对具有高社会地位的群体时产生更多的自我欺骗，社会奖励的存在也会使个体产生较多的自我欺骗(Ren et al., 2018; Sloman et al., 2010)。但有研究发现当个体得到真实或精确的反馈之后，或个体的认知意识降低，往往会降低自我欺骗的发生(Liu et al., 2019; 钟罗金 等, 2019)。而情感效价在自我欺骗的作用主要体现在根据个体的关心程度有选择地评估信息(Galeotti, 2016; Lauria & Preissmann, 2018)。处于消极情绪中的个体会通过自我欺骗以达到调节心理状态的目的；另外个体会在积极情绪状态下产生更多的自我欺骗以达到继续维持积极正面的形象((Lauria & Preissmann, 2018; 谭成慧 等, 2020)。总之，以往的研究表明个体在接受模糊信息之后往往产生较多的自我欺骗，而接受信息逐渐明确的个体往往会降低自我欺骗的出现。但这种现象在各种情境中是稳定存在的吗？信息的传递并非只有简单内容的表达，个体自我欺骗的增多或减少是否会受到情感效价的影响？还不得而知。

以往学者大多采用量表或行为实验展开对自我欺骗的研究，随着认知神经科学的发展，有研究将 fMRI 和 TMS 等脑科学技术运用于自我欺骗的探索中(Duran et al., 2020; Farrow et al., 2015; 朱晶, 蒋京川, 2014)。有研究发现 N1 是视觉感知刺激比较敏感的部位，自我欺骗出现在反应时的早期阶段，主要反映在个体大脑感知觉的敏感部位(Jian et al., 2019)。P2 成分出现在早期认知阶段，负性刺激往往比正性刺激占据更多的认知资源，峰值潜伏期在 100~200 ms 之间的 P2 对所呈现刺激的效价比较敏感 (Chen & Chai, 2010; Lai & Huettig, 2016; Paulmann & Kotz, 2008)。Kutas 和 Hillyard (1980) 发现潜伏期为 300~600 ms 的 N400 是研究语言加工时常用的 ERP 负性成分。前人研究发现当启动刺激与目标刺激语义相关时，诱发的 N400 成分较低；而当启动刺激与目标刺激语义违反时诱发的 N400 成分较高(Kutas & Federmeier, 2011)。N400 成分通常被认为是关于语义的电生理学指标，但是 N400 不仅能够反映个体对语境整合的难易程度，还会受到情绪信息的影响(Federmeier & Kutas, 2001)，因此，本研究选择 N1、P2 和 N400 成分作为探索自我欺骗内在神经机制的敏感指标。

综上所述，自我欺骗作为人类行为中一种常见的现象，由于其操作诱发和检测的复杂性导致难以进行大量的实验室研究和无法广泛地探究其背后的神经机制。以往学者从单一因素的角度对自我欺骗进行了研究，但是却得到极具差异性的结果(Chance et al., 2011; Chance & Norton, 2015; Sloman et al., 2010)。这些不一致的结论无法使学者们在对自我欺骗的研究中得到统一的参照。此外，过于单一的变量使我们无法从整体角度理解个体对自我欺骗的加工。因此，自我欺骗的研究者不仅需要寻求稳定的研究结果，而且也需要不断采用更加全面整体的研究角度展开对自我欺骗的探索。因此，本研究提出问题 1：以往学者只采用单一模糊反馈信息对自我欺骗进行研究，但也有研究发现自我欺骗是个体为了避免伤害性事件带给自己打击而采取的一种情感应对方式，情感效价在自我欺骗中发挥着重要的作用(Lauria & Preissmann, 2018; Duran et al., 2020; 朱艳丽 等, 2016)，但是模糊反馈和效价以怎么样的机制共同影响自我欺骗的产生目前还不得而知。为解决问题 1，本研究设计实验 1 以探索不同反馈效价的模糊信息对自我欺骗的影响，研究假设相比于消极模糊反馈的条件，个体在积极模糊反馈中产生更多的自我欺骗。同时提出问题 2：反馈信息的多样性可能使个体在实验 1 中采用模糊信息得到的结果无法推广到非模糊信息的背景中。为解决这个问题，本研究设计实验 2 将模糊反馈信息进行精确化；并保留反馈效价在自我欺骗的作用。研究假设相比于消极精确反馈的条件，个体在积极精确反馈中也产生较多的自我欺骗。本研究进一步提出问题 3：前两个实验分别考察了模糊效价反馈和精确效价反馈对自我欺骗产生的作用，但是反馈效价和反馈类型是否会整体对自我欺骗产生交互作用，目前还不得而知。为了更加全面和整体对自我欺骗进行探索，本研究设计实验 3 探索反馈类型和反馈效价对自我欺骗的整合作用，以及采用 ERP 技术探索不同条件对自我欺骗产生差异的背后机制(刘燕妮, 舒华, 2003)。

2　实验 1：不同效价的模糊反馈对自我欺骗的影响

2.1　研究目的和假设

实验 1 探索不同效价的模糊反馈促进自我欺骗的产生。研究假设在测试 1 有答案提示的影响下，被试在测试 2 第一阶段的预测与真实分数之差较大；相比于消极模糊反馈，积极模糊反馈条件的被试提高测试 2 第二阶段的预测分数。

2.2　方法

2.2.1　被试

使用 G-Power 3.1 (Faul et al., 2009)进行的功效分析表明，在保证得到效应量 f = 0.25 (中)的前提下，设定 $\alpha = 0.05$，至少需要 44 名被试才能达到 95% $(1 - \beta)$的统计检验力。最后招募 70 名湖南师范大学在校大学生，年龄在 18~24 岁之间，平均年龄为 20.58 (SD = 1.54)，男女比例均衡。视力、矫正视力均正常，所有被试之前均未参加过类似实验。本实验获得湖南师范大学伦理委员会的认可，并且被试签署实验知情同意书，在实验后给予被试一定的报酬。

2.2.2　实验设计

采用 3 (模糊反馈效价：积极 vs.消极 vs.无) × 2 (组别：有答案提示组 vs.无答案提示组)混合实验设计。模糊反馈效价为被试内变量，因变量为测试 2 第一阶段的预测分数与真实分数之差(预测分数：被试在完成测试 1 之后预测将要完成测试 2“红点图”的分数；真实分数：被试真正完成测试 2“红点图”的分数)；测试 2 第二阶段的预测分数(被试在测试 2 接受不同反馈效价之后对“预测题目”的预测分数)。

2.2.3　实验材料

(1)自我欺骗的测量——前瞻范式(Chance et al., 2011; Liu et al., 2019)。该范式是由 150 张红点图组成的两次测试(测试 1 和测试 2)。在测试 1 阶段(见图 1A)，相同 60 张红点图分别做成有答案标识和没有答案标识的两类。测试 1 阶段，被试的任务是判断被一条对角线一分为二的红点图，哪边的红点多，被试按“J”键表示“右边”的点数多；被试按“F”键则表示判断“左边”的点数多。测试 2 阶段(见图 1B)具有 90 张没有答案标识的红点图，被试不仅需要对红点图进行判断，还需要在接受反馈条件后对屏幕中出现的“预测题目”进行回答并在电脑上输入预测分数的数字值。

(2)反馈效价的类型。本实验采用统一的信息来表达积极、消极和无效价模糊反馈(见图 1C)。由于反馈效价是本实验结合以往研究进行改编，所以我们在开始实验之前找 60 名大学生对 3 种反馈效价进行测量。通过单因素方差分析的计算可知，$F(2, 58) = 6.23$, $p = 0.03$, $\eta_p^2 = 0.08$。事后比较发现，积极模糊反馈与消极模糊反馈具有显著差异，$p < 0.05$；与无反馈具有显著差异，$p < 0.05$；消极模糊反馈与

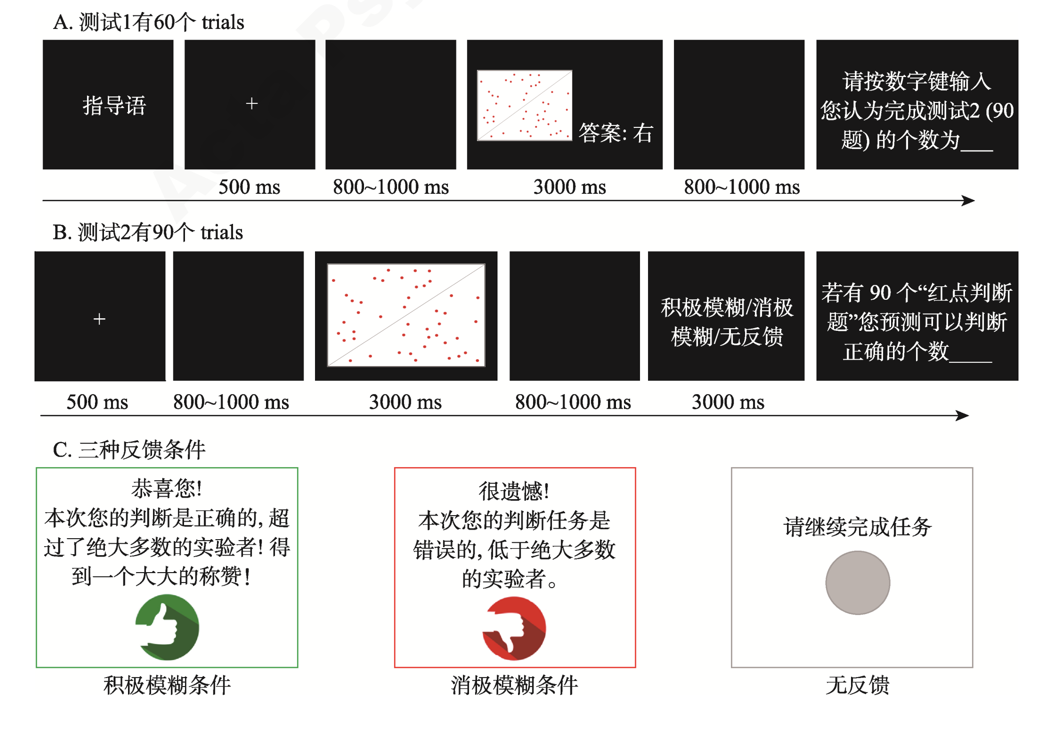


图 1　实验 1 的实验流程图。A.测试 1 的实验流程图; B.测试 2 的实验流程图; C. 3 种反馈条件

无反馈具有显著差异, $p < 0.05$。

(3)自我欺骗的量表。本实验采用具有 20 道题目的自我欺骗量表, 它的一致性系数是 0.82 (Egerton et al., 2010; Gur & Sackeim, 1979), 题目中的偶数题目需要反向计分。

2.2.4 实验程序

被试在这个实验中共完成 2 个测试。在测试 1 中, 被试先阅读指导语, 然后是 500 ms 的"+", 800~1000 ms 的黑屏, 3000 ms 的红点图判断(答案提示组看到是右下角具有"答案: 右"的红点图, 无答案提示组看到是没有提示的红点图), 800~1000 ms 的黑屏, 然后是下一个 trial, 测试 1 共有 60 个 trials。完成测试 1 后, 被试填写将要完成测试 2 的预测分数(见图 1A)。在测试 2 中, 500 ms 的"+", 800~1000 ms 的黑屏, 3000 ms 的红点图判断, 800~1000 ms 的黑屏, 3000 ms 的随机出现 3 种模糊反馈(积极、消极和无), 测试 2 要求被试在接受每一个反馈后都对题目进行预测判断(预测判断是让被试根据对预测题目的回答输入预测数值) (见图 1B), 测试 2 共有 90 个 trials。

2.2.5 数据统计与分析

根据研究目的和假设, 本实验采用 SPSS 21.0 对自我欺骗量表和被试对测试 2 第一阶段的预测分数与真实分数之差进行独立样本 t 检验, 然后对测试 2 第二阶段的预测答题成绩进行两因素混合重复方差分析。

2.3 研究结果

2.3.1 操作性检验

(1)自我欺骗量表

通过独立样本 t 检验可知, $t(1) = -0.76$, $p = 0.45$, Cohen's $d = 1.07$, 95% CI = [−8.51, 3.85]。无答案提示组(80.87 ± 11.82)和有答案提示组(83.20 ± 12.10)两组被试在自我欺骗量表的得分没有达到显著差异。

(2)测试 2 第一阶段的结果比较

通过独立样本 t 检验可知, 在测试 2 第一阶段的预测分数上两组被试达到极显著差异, $t(1) = -4.83$, $p < 0.001$, Cohen's $d = 6.85$, 95% CI = [−24.03, −9.96]; 相比于无答案提示组, 有答案提示组具有较高的预测分数。通过独立样本 t 检验可知, 在测试 2 第一阶段的真实分数上两组被试没有达到显著差异, $p > 0.05$。通过独立样本 t 检验可知, 在测试 2 第一阶段的预测与真实分数之差上两组被试达到显著差异, $t(1) = -2.10$, $p = 0.04$, Cohen's $d = 3.14$, 95% CI = [−20.53, −5.09]; 相比于无答案提示组, 有答案提示组的被试具有较高的预测与真实分数之差(自我欺骗产生) (见表 1)。

表 1 无答案提示组和有答案提示组的被试在测试 2 第一阶段的得分($M \pm SD$)

分数	$M \pm SD$	t	p
测试 2 的预测分数			
无答案提示组	48.36 ± 13.34	-4.83^{***}	0.00
有答案提示组	65.36 ± 13.88		
测试 2 的真实分数			
无答案提示组	53.82 ± 6.95	−0.22	0.83
有答案提示组	54.26 ± 8.91		
测试 2 的预测与真实分数之差			
无答案提示组	−4.95 ± 14.85	-2.10^{*}	0.04
有答案提示组	9.80 ± 16.95		

注: 每组 $n = 35$

2.3.2 测试 2 第二阶段的预测分数

经过重复测量方差分析可知, 组别的主效应显著, $F(1, 68) = 5.29$, $p = 0.03$, $\eta_p^2 = 0.08$, 95% CI = [−13.71, −0.95]; 有答案提示组在测试 2 第二阶段的预测分数显著大于无答案提示组。模糊反馈效价的主效应极显著, $F(2, 68) = 29.55$, $p < 0.001$, $\eta_p^2 = 0.34$。事后比较发现: 在积极反馈下, 被试在测试 2 第二阶段的预测分数显著大于消极反馈和无反馈条件下的预测分数, $ps < 0.001$, 95% CI = [2.68, 5.96], 95% CI = [1.27, 3.53]; 无反馈显著大于消极反馈条件, $p = 0.003$, 95% CI = [0.57, 3.26]。组别和模糊反馈不具有交互作用, $F(2, 68) = 0.09$, $p = 0.91$ (见图 2)。

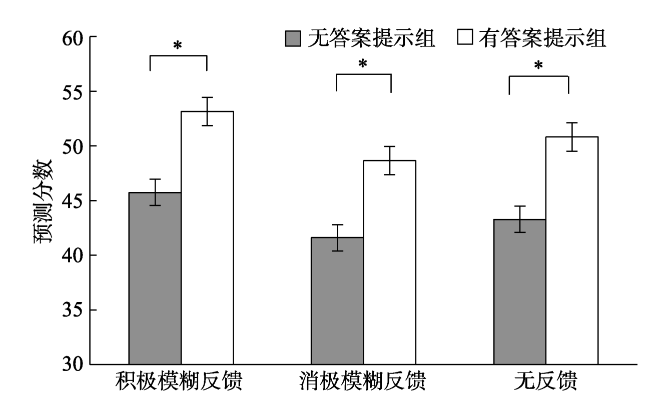


图 2 测试 2 第二阶段不同效价模糊反馈的预测分数; 误差线表示标准误,
$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$, 下同.

2.4 讨论

研究结果发现在积极模糊反馈下所形成的自我欺骗多于无反馈条件, 这与 Sloman 等人所提出模糊反馈对自我欺骗的产生有显著影响的结果相

一致(Liu et al., 2019; Sloman et al., 2010)。另外, 研究结果也发现个体的自我欺骗会受到积极反馈效价的影响, 这表明了情感效价在自我欺骗中具有促进作用(Lauria & Preissmann, 2018)。实验结果较好的回答了研究问题 1, 即积极模糊的条件促进个体产生更多的自我欺骗。遗憾的是, 目前尚未查到其他的研究探索了反馈效价在自我欺骗中的作用, 即探索积极效价对自我欺骗的影响是否具有稳定性。

本实验希望在下一个实验中继续添加不同反馈效价的信息, 再次探讨反馈效价等因素对自我欺骗的影响。在日常生活中, 个体不仅接受模糊的反馈, 而且会受到精确与清晰明了等信息的影响。以往有研究用金钱作为精确信息展开对自我欺骗的探索, 但结果发现精确的金钱反馈不会对自我欺骗产生影响(Chance et al., 2011; Chance & Norton, 2015)。在此基础上, 本研究想进一步探索不同效价的精确反馈对自我欺骗的影响, 即个体在完成任务后可以获得精确金钱的报酬, 那么这个不断累加的精确金钱是否会对自我欺骗产生影响吗？这个问题将在实验 2 中进行验证。

3　实验 2：不同效价的精确反馈对自我欺骗的影响

3.1　研究目的和假设

实验 2 探索反馈的精确化是否会削弱自我欺骗的发生。研究假设：相比于消极反馈, 被试在积极精确和无反馈条件下产生较多的预测分数。相比于消极精确的条件, 有答案提示组的被试在积极精确和无反馈条件下在测试 2 第二阶段的预测分数都较低。

3.2　方法

3.2.1　被试

使用 G-Power 3.1 进行的功效分析表明, 在保证得到效应量 f = 0.25(中)的前提下, 设定 α = 0.05, 至少需要 44 名被试才能达到 95% ($1-\beta$)的统计检验力。最终招募 72 名湖南师范大学的在校大学生, 年龄在 18~24 岁之间, 平均年龄为 20.15 岁(SD = 2.29)。本实验删除 6 名具有极端数据的被试(1 名被试无规则的按键、通过 SPSS 软件计算有 3 名被试的数据离群严重以及 2 名没有完整做实验的被试), 最后有 66 个被试统计到数据分析中(男生有 32 人)。所有被试的视力、矫正视力属于正常; 均未参加过类似实验。本实验获得湖南师范大学伦理委员会的认可, 被试签署知情同意书和在实验后给予一定的报酬。

3.2.2　实验设计

本实验采用 3 (精确反馈效价：积极 vs.消极 vs.无) × 2 (组别：有答案提示组 vs.无答案提示组)两因素混合实验设计。被试内变量为精确反馈效价, 因变量为测试 2 第一阶段的预测分数与真实分数之差和测试 2 第二阶段的预测分数。

3.2.3　实验材料

(1)自我欺骗的前瞻范式和量表, 同实验 1。

(2)反馈效价的类型。实验 2 采用精确反馈效价(积极、消极和无)的实验材料, 告诉被试在正式做实验任务之前他们有 10 元的基础费, 在完成进行实验任务中会根据实验反馈获得或减掉一定的金额(见图 3C)。本实验材料是结合以往研究进行改编, 在正式使用材料前找 67 名大学生对 3 种精确反馈效价进行测量。通过单因素方差分析可知, 结果发现实验材料具有主效应, $F(2, 65) = 11.57$, $p = 0.05$, $\eta_p^2 = 0.12$。事后比较发现, 积极精确反馈与消极精确反馈、无反馈具有显著差异, $ps < 0.05$; 消极精确反馈与无反馈具有显著差异, $p < 0.05$。

3.2.4　实验程序

同实验 1, 唯一不同之处在于被试接受不同效价的精确反馈信息(见图 3)。

3.2.5　数据统计与分析

根据研究目的和假设, 本实验采用 SPSS 21.0 对自我欺骗量表和测试 2 第一阶段的预测与真实分数之差进行独立样本 t 检验; 对测试 2 第二阶段的预测分数进行两因素混合重复方差分析。

3.3　研究结果

3.3.1　操作性检验

(1)自我欺骗量表的得分

通过独立样本 t 检验可知, $t(1) = -1.09$, $p = 0.28$, Cohen's d = 1.49, 95% CI = [−7.13, 2.28]; 无答案提示组(80.32 ± 8.06)和有答案提示组(82.96 ± 8.54)两组被试在量表上的得分没有达到显著差异。

(2)两组被试在测试 2 第一阶段的结果比较

通过独立样本 t 检验可知, 两组被试在测试 2 第一阶段的预测分数上达到显著差异, $t(1) = -2.89$, $p = 0.01$, Cohen's d = 4.27, 95% CI = [−19.72, −3.55]; 相比于无答案提示组, 有答案提示组的被试具有较高的预测分数。通过独立样本 t 检验可知, 两组被试在测试 2 第一阶段的预测分数与真实分数之差上具有显著的差异, $t(1) = -2.12$, $p = 0.04$, Cohen's d = 3.05, 95% CI = [−20.51, −0.57]; 相比于无答案提示

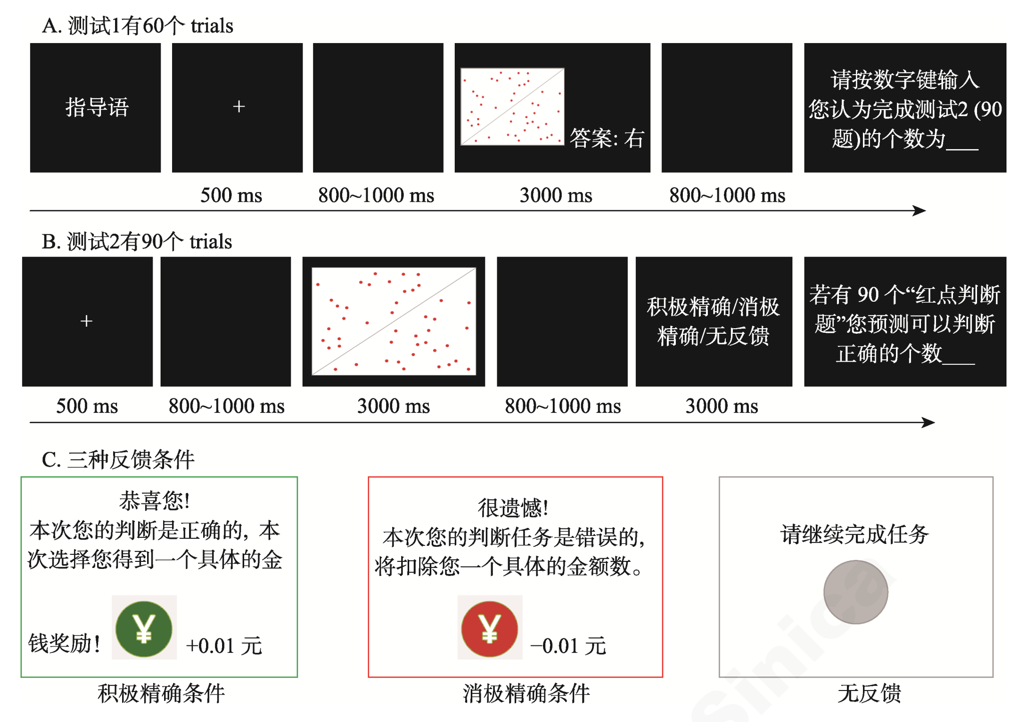


图 3　实验 2 的实验流程图。A.测试 1 的实验流程图; B.测试 2 的实验流程图; C. 3 种反馈条件

组, 有答案提示组的被试具有较高的差值。通过独立样本 t 检验可知, 两组被试在测试 2 的真实分数上没有达到显著差异, $p > 0.05$ (见表 2)。

表 2　无答案提示组和有答案提示组的被试在测试 2 第一阶段的得分($M \pm SD$)

分数	$M \pm SD$	t	p
测试 2 的预测分数			
无答案提示组	52.93 ± 16.59	−2.89*	0.006
有答案提示组	64.57 ± 13.66		
测试 2 的真实分数			
无答案提示组	53.72 ± 7.53	−0.52	0.61
有答案提示组	54.82 ± 8.50		
测试 2 的预测与真实分数之差			
无答案提示组	0.79 ± 19.57	−2.12*	0.038
有答案提示组	9.75 ± 147.91		

3.3.2　测试 2 第二阶段的预测分数

经过重复测量方差分析可知, 组别的主效应不显著, $F(1, 64) = 1.63$, $p = 0.21$, 95% CI = [−1.87, 8.46]。精确反馈的主效应显著, $F(2, 64) = 20.70$, $p < 0.001$, $\eta_p^2 = 0.27$, 事后比较发现: 在积极精确反馈下, 被试的测试 2 第二阶段的预测分数极显著大于消极反馈和无反馈条件的预测分数, $p < 0.001$, 95% CI = [2.55, 7.13]; $p < 0.001$, 95% CI = [0.84, 2.68]; 无反馈条件下测试 2 第二阶段的预测分数显著大于消极精确反馈, $p < 0.001$, 95% CI = [0.96, 5.20]。

组别和不同效价的精确反馈存在交互作用, $F(2, 64) = 4.01$, $p = 0.02$, $\eta_p^2 = 0.06$。进一步的简单效应分析可知, 在消极精确反馈条件下, $p = 0.03$, 95% CI = [38.53, 45.76], 无答案提示组在测试 2 第二阶段的预测分数显著大于有答案提示组; 在无反馈条件和积极精确反馈条件下, 无答案提示组在测试 2 第二阶段的预测分数没有显著大于有答案提示组, $p = 0.32$; $p = 0.61$ (见图 4)。

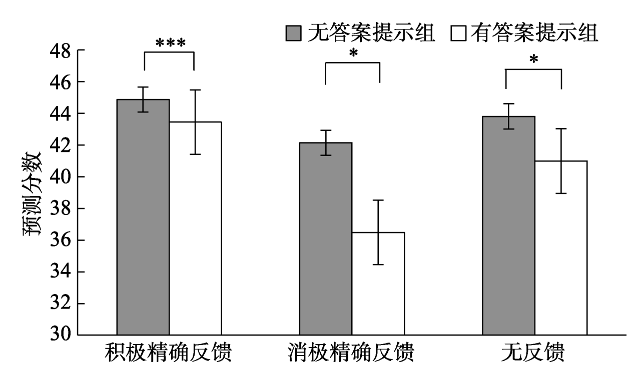


图 4　两组被试在测试 2 中不同效价精确反馈条件下的预测分数

3.4　讨论

研究结果发现有答案提示组的个体在测试 2 第二阶段中没有获得较高的预测分数, 该结果与实验 1 具有相反的表现, 其原因可能是因为个体接受不同的实验刺激, 精确反馈信息不能让被试继续维持高信心预测将来自己获得较高的成绩(钟罗金, 汝

涛涛, 2021)。另外这也可能说明, 有答案提示组的个体可以通过具体金钱的反馈认识自己的能力和对自我具有清晰的认知, 理解了在测试 1 所得到的好成绩是因为有外在答案信息的提示所致。个体对自己接下来的反应比较谨慎和保守, 所以在测试 2 中有答案提示组的个体具有较低的预测分数, 精确信息的反馈可能削弱了自我欺骗的产生(Chance et al., 2015; Liu et al., 2019)。实验 2 是对研究问题 2 的解答, 实验结果可能表明积极信息可以促进自我欺骗的产生, 但消极精确反馈信息降低自我欺骗的产生。

实验 1 和实验 2 发现不同效价的模糊和精确反馈对自我欺骗产生了巨大相反的作用。产生该结果的原因也有可能是因为不同实验条件和不同批次的被试等原因而造成的。对此, 研究者希望在接下来的实验中弥补这些研究的缺点, 从整体和全面的角度对自我欺骗进行科学研究。在实验 3 中, 借助 ERP 技术探索反馈类型和反馈效价对自我欺骗产生差异的内在神经机制, 为自我欺骗领域提供新的研究视角。

4　实验 3：探索反馈类型和反馈效价影响自我欺骗产生差异的内在神经机制

4.1　研究目的和假设

相比于在积极精确反馈下, 被试在积极模糊条件下产生更多的自我欺骗次数。相对于精确反馈, 被试在模糊反馈条件诱发更大的 P2 以及更小的 N1、N400 成分; N400 成分是测量自我欺骗内在神经机制的有效指标。

4.2　研究方法

4.2.1　被试

使用 G-Power 3.1 (Faul et al., 2009)进行的功效分析表明, 在保证得到效应量 f = 0.25 (中)的前提下, 设定 α = 0.05, 至少需要 34 名被试才能达到 80% (1 – β)的统计检验力。最终招募 41 名湖南师范大学的在校大学生, 删除 6 名具有极端数据的被试(3 名被试因过度的眼动等伪迹被排除在分析之外、删除 3 名通过 SPSS 软件计算数据离群非常严重的被试), 最后有 35 个被试统计到数据分析中。年龄在 18~22 岁之间, 平均年龄为 19.90 岁(SD = 1.45), 男女比例均衡。被试的视力或矫正视力属于正常, 之前均未参加过类似实验。本研究获得湖南师范大学伦理委员会的认可, 被试签署知情同意书和给予一定的实验报酬。

4.2.2　实验设计

实验采用 2 (反馈类型：模糊 vs.精确) × 2 (反馈效价：积极 vs.消极)两因素被试内实验设计。自变量为反馈类型和反馈效价。因变量为测试 2 第二阶段自我欺骗的次数, 以及 ERP 数据的 N1、P2 和 N400 成分。

4.2.3　实验材料

自我欺骗的量表, 同实验 1。

自我欺骗的测量——前瞻范式, 同实验 1。

反馈类型和反馈效价的材料。实验 3 选择实验 1 中的积极和消极模糊反馈以及实验 2 中的积极精确和消极精确反馈组成反馈类型和反馈效价的测试。实验 3 将 4 种条件设置成被试内实验设计(2 个 blocks)。本研究在实验程序进行 4 种实验条件(积极和消极模糊、积极和消极精确)的平衡, 主要体现在 2 个 blocks 之间的平衡以及一个 block 中 trials 的随机(见图 5 C)。

4.2.4　实验程序

首先, 在实验 3 中被试完成带有答案提示的测试 1 (见图 5A), 然后完成测试 2 没有答案提示的 240 个红点图(见图 5B)。本实验的随机主要体现在实验操作中, 一半的被试先完成模糊反馈, 然后完成精确反馈; 另一半被试则相反。被试的任务除了完成红点图的判断之外。需要在每次随机接受某种反馈条件之后, 对预测题目的三分之二分值进行选择(“大于”代表自我欺骗的出现)。

4.2.5　脑电记录与分析

使用国际 10-20 系统扩展 64 导脑电放大器(ANT Neuro, Enschede, Netherlands)采集脑电和眼电数据, 在线记录时以 CPz 点为参考电极, 离线转为双侧乳突平均参考。同时记录双眼外侧的水平眼电(HEOG)和右眼上下眶的垂直眼电(VEOG), 采样频率为 500 Hz/导, 所有电极与头皮之间阻抗都小于 5 kΩ, EEG 数据分析采用 EEGLAB 工具包(Delorme & Makeig, 2004)。首先, 对数据进行滤波, 滤波工具是 EEGLAB 工具包内置的 Hamming windowed sinc FIR (finite impulse response)滤波器, 参数为 0.1~30 Hz (filter slopes: 24 dB/octave); 然后, 利用 ICA (独立成分分析, Independent Component Analysis)方法去除脑电中的水平和垂直眼电和伪迹 (Delorme & Makeig, 2004; Plöchl et al., 2012); 之后, 对数据进行分段, 分段时间为−200 ms (作为基

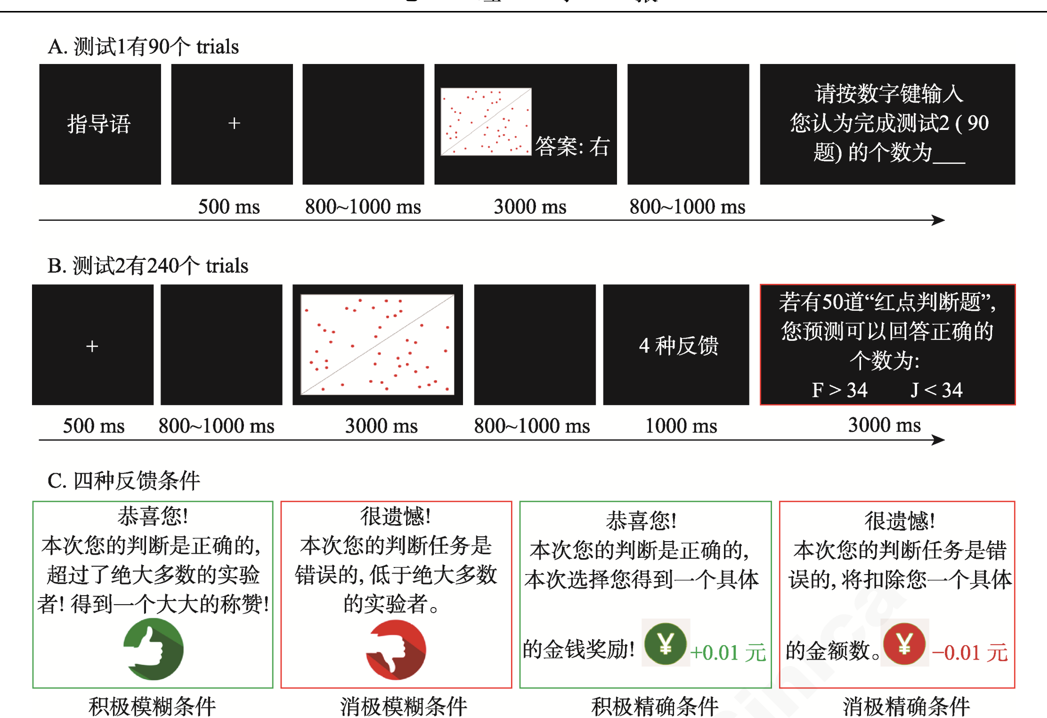


图 5 A.实验 3 的测试 1 流程图; B.实验 3 的测试 2 流程图; C. 4 种反馈条件

注：在 B 部分"若有 50 道'红点判断题'，您预测可以回答正确的个数为'F > 34; J < 34'"，其中"50"这个数字的选择是根据实验 1 和实验 2 在测试 2 进行预测判断的题目而来，而"34" (如预测题目"50"的三分之二分值)这个数字的选择是根据实验 1 和实验 2 被试对测试 2 第一阶段的预测分数进行计算而来，因为前两个实验都发现被试对 90 个题目的预测大约在"答对 60 道题目"。红框表示记录脑电成分的界面。

线矫正)至 800 ms; 最后排除其他波幅大于±70 μV 的伪迹信号。本研究在分析有效的脑电波时，只对不同反馈条件出现后被试进行决策前的−200 ms 到 800 ms 进行分析，在分析脑电波时，把前 200 ms 设为基线，然后求各个反馈条件下的平均脑电波幅。

4.2.6 数据统计与分析

本实验采用 SPSS 21.0 对不同反馈条件下产生自我欺骗的次数进行重复方差分析。根据本研究目的、脑地形图及视觉检测，我们分析的是时间窗 N1 (0~50 ms)、P2 (100~250 ms)和 N400 (300~500 ms)的平均成分，进行多因素被试内方差分析 2 (反馈类型：模糊 vs.精确) × 2 (反馈效价：积极 vs.消极) × 5 [脑区：额区(F3, Fz, F4) vs. 额中区(FC3, FCz, FC4) vs. 中央区(C3, Cz, C4) vs. 中顶区(CP3, CPz, CP4) vs. 顶区(P3, Pz, P4)]；以及 2 (反馈类型：模糊 vs. 精确) × 2 (反馈效价：积极 vs. 消极) × 3 [脑半球：左半球(F3、FC3、C3、CP3、P3) vs. 中央区(Fz、FCz、Cz、CPz、Pz) vs. 右半球(F4、FC4、C4、CP4、P4)]。所有主效应和交互作用的 p 值均采用 Greenhouse-Geisser 法校正，事后配对比较采用 Bonferroni 法校正，数据分析采用 SPSS 21.0 (IBM Corp., Armonk, NY, USA)进行分析。

4.3 实验结果

4.3.1 操作性检验

(1) 3 个实验(有答案提示组)在自我欺骗量表的对比

通过单因素方差分析可知，$F(2, 34) = 0.23$, $p = 0.80$, $\eta_p^2 = 0.01$，3 组被试在量表得分上无显著差异(见表 3)。

表 3 三个实验(有答案提示组的被试)在自我欺骗量表上的得分($M \pm SD$)

组别	N	$M \pm SD$	F	p
实验 1 的有答案提示组	35	83.20 ± 12.10		
实验 2 的有答案提示组	33	81.77 ± 11.72	0.23	0.80
实验 3 的有答案提示组	35	81.40 ± 11.45		

(2) 3 个实验中有无答案提示组的被试在测试 2 第一阶段的结果比较

通过单因素方差分析可知，3 组被试在测试 2 第一阶段的预测分数的主效应显著，$F(2, 34) = 12.50$, $p < 0.001$, $\eta_p^2 = 0.22$，事后比较发现：实验 3 有答案提示组被试的预测分数显著大于实验 1 的无答案提示组和实验 2 的无答案提示组，$p < 0.001$, 95% CI = [−26.49, −11.15]; $p < 0.001$, 95% CI = [−20.44, −5.23] (见表 4)。

表 4　三个实验中有无答案提示组的被试在测试 2 第一阶段的结果比较

组别	N	$M \pm SD$	F	p
实验 1 的无答案提示组	35	48.36 ± 13.34		
实验 2 的无答案提示组	33	52.93 ± 16.59	12.50***	0.00
实验 3 的有答案提示组	35	81.40 ± 11.45		

注：实验 1 和实验 2 的无答案提示组与实验 3 的有答案提示组形成不同的组别，分析 3 组被试在测试 2 第一阶段的结果，以确定实验 3 有答案提示组的被试产生自我欺骗。

4.3.2　测试 2 第二阶段不同反馈条件下自我欺骗出现的次数

通过重复测量方差分析可知，反馈类型具有主效应, $F(1, 34) = 4.24$, $p = 0.05$, $\eta_p^2 = 0.14$, 95% CI = [0.03, 15.69]；在模糊反馈条件下出现自我欺骗的次数显著大于精确反馈条件的次数。反馈效价具有主效应, $F(1, 34) = 6.08$, $p = 0.02$, $\eta_p^2 = 0.18$, 95% CI = [0.52 5.76]；在积极效价条件中出现自我欺骗的次数显著大于消极效价的次数。反馈类型和反馈效价存在交互作用, $F(3, 34) = 9.76$, $p < 0.001$, $\eta_p^2 = 0.27$。经过简单效应分析可知，在积极反馈效价下，相比于精确反馈类型，模糊反馈类型具有更多次数的自我欺骗, $p < 0.001$, 95% CI = [32.66, 44.12]；在消极反馈效价下，两种反馈类型在自我欺骗出现的次数上无显著差异, $p = 0.459$ (见图 6)。

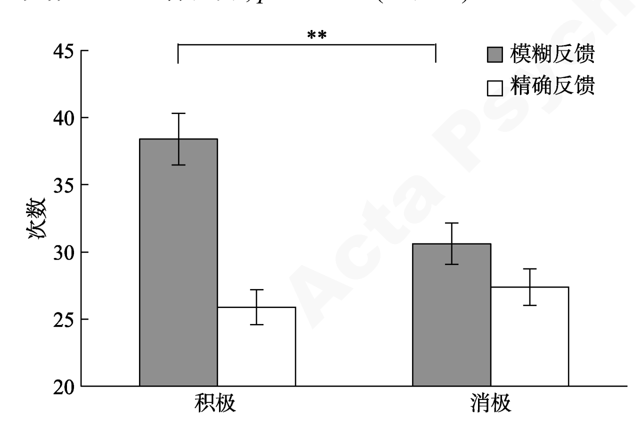


图 6　四种反馈条件下自我欺骗出现的次数

4.3.3　测试 2 第二阶段不同反馈条件下自我欺骗出现的次数(反应时)

通过重复测量方差分析可知，反馈类型的主效应有边缘显著, $F(1, 34) = 2.99$, $p = 0.09$, $\eta_p^2 = 0.10$, 95% CI = [819, 1186]；在模糊反馈条件下出现自我欺骗的反应时显著大于精确反馈条件。反馈效价的主效应不显著, $F(1, 34) = 0.89$, $p = 0.35$。反馈类型和反馈效价不存在交互作用, $F(3, 34) = 1.39$, $p = 0.25$ (见图 7)。

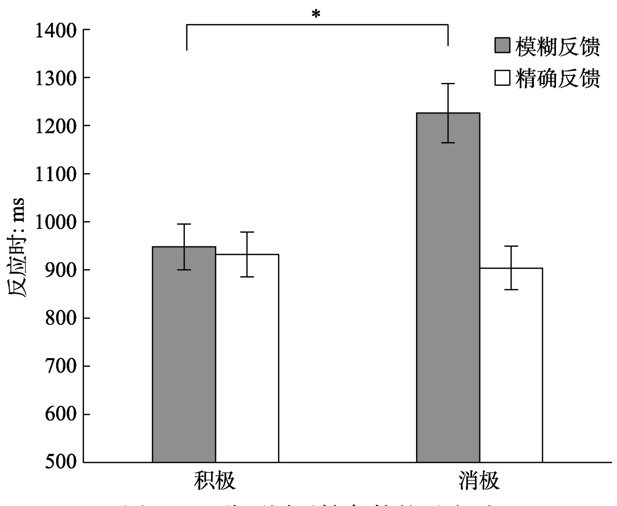


图 7　四种不同反馈条件的反应时

4.3.4　ERPs 结果

(1) N1 成分

通过对 N1 成分进行 2 (反馈类型：模糊 vs.精确) × 2 (反馈效价：积极 vs.消极) × 5 (脑区：额区 vs.额中区 vs.中央区 vs.中顶区 vs.顶区)的重复测量方差分析可知，反馈类型的主效应边缘显著, $F(1, 34) = 3.18$, $p = 0.09$, $\eta_p^2 = 0.11$, 95% CI = [−3.39, −0.24]；在模糊反馈条件下出现自我欺骗的 N1 (−1.051 μV)显著小于精确反馈条件(0.521 μV)。反馈效价的主效应不显著, $F(1, 34) = 0.99$, $p = 0.33$。脑区的主效应也不显著, $F(4, 34) = 0.82$, $p = 0.37$。反馈类型和反馈效价不存在交互作用, $F(3, 34) = 0.18$, $p = 0.68$。反馈类型和脑区不存在交互作用, $F(9, 34) = 1.25$, $p = 0.27$。反馈效价和脑区也不存在交互作用, $F(9, 34) = 0.81$, $p = 0.78$。

反馈类型、反馈效价和脑区三者存在交互作用, $F(19, 34) = 6.13$, $p = 0.02$, $\eta_p^2 = 0.19$。简单简单效应分析发现，在积极反馈条件下，相比于精确反馈，模糊反馈在额区诱发显著的 N1 成分, $F(9, 34) = 7.97$, $p = 0.01$, 95% CI = [−2.91, −0.62]；在额中区诱发显著的 N1 成分, $F(9, 34) = 5.05$, $p = 0.03$, 95% CI = [−2.85, −0.86]；在中央区诱发显著的 N1 成分, $F(9, 34) = 4.68$, $p = 0.04$, 95% CI = [−2.58, −0.91]；在中顶区诱发显著的 N1 成分, $F(9, 34) = 3.80$, $p = 0.06$, 95% CI = [−2.45, −0.69]；其他的条件下，均为无显著差异, $p > 0.05$ (见图 8)。

通过对 N1 成分进行 2 (反馈类型：模糊 vs.精确) × 2 (反馈效价：积极 vs.消极) × 3 (脑半球：左半球 vs.中央区 vs.右半球)的重复测量方差分析发现，反馈类型的主效应边缘显著, $F(1, 34) = 3.10$, $p = 0.09$, $\eta_p^2 = 0.09$, 95% CI = [−2.99, −0.23]；在模糊

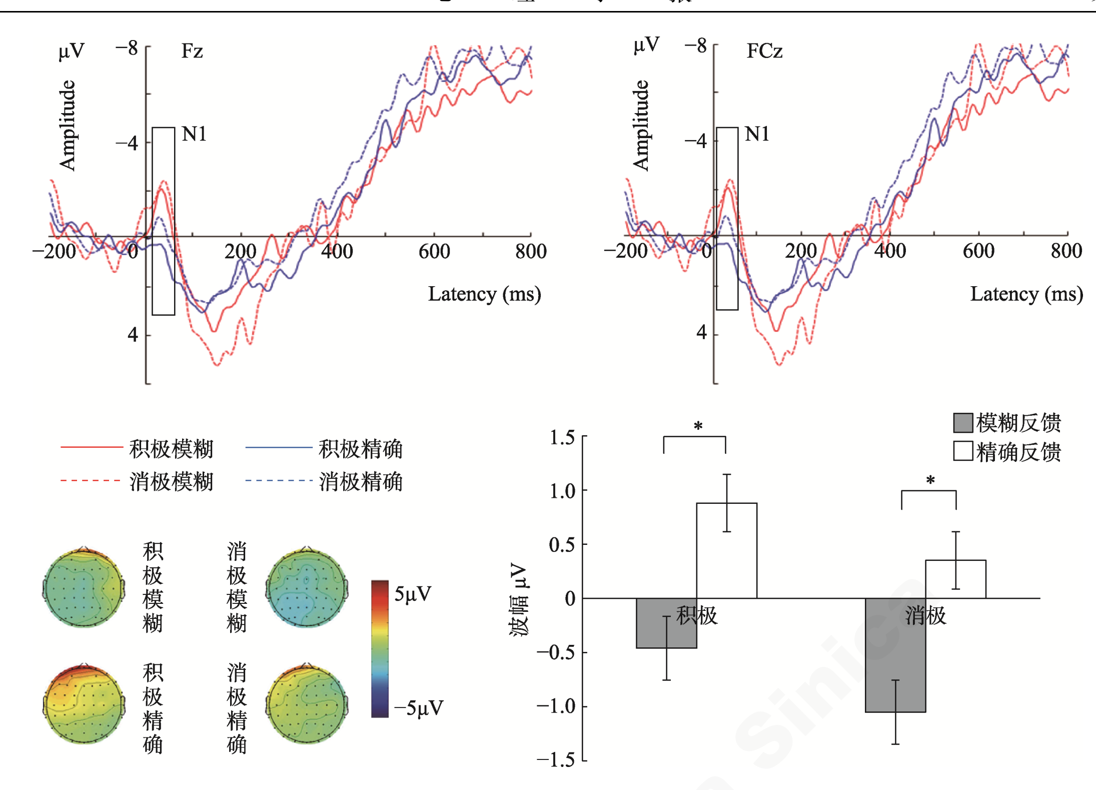


图 8　四个条件在 Fz 和 FCz 电极点的 N1 (时间窗：0~50 ms)的 ERP 总平均图、脑地形图和所有分析点的 N1 成分平均值柱状图, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

反馈条件下出现自我欺骗的 N1 (−1.76 μV)显著小于精确反馈条件(0.63 μV)。反馈效价的主效应不显著, $F(1, 34) = 2.57$, $p = 0.12$。脑半球的主效应不显著, $F(2, 34) = 0.87$, $p = 0.77$。反馈类型和反馈效价不存在交互作用, $F(3, 34) = 0.02$, $p = 0.90$。

反馈类型和脑半球存在交互作用, $F(5, 34) = 2.65$, $p = 0.08$, $\eta_p^2 = 0.08$; 简单效应分析可知, 在左半球, 相比于精确反馈的条件(0.98 μV), 模糊反馈(−0.84 μV)在诱发 N1 成分上具有显著差异, $F(1, 34) = 6.44$, $p = 0.02$, 95% CI = [−2.21, −0.53]; 在中央区, 相比于精确反馈(0.53 μV), 模糊反馈(−1.03 μV)在诱发 N1 成分上具有显著差异, $F(1, 34) = 3.02$, $p = 0.09$, 95% CI = [−2.76, −0.71]; 在右半球, 相比于精确反馈, 模糊反馈不显著, $F(1, 34) = 0.79$, $p = 0.38$。反馈效价和脑半球不存在交互作用, $F(5, 34) = 0.05$, $p = 0.95$。反馈类型、反馈效价和脑半球不存在交互作用, $F(19, 34) = 0.13$, $p = 0.88$。

(2) P2 成分

通过对 P2 成分进行 2 (反馈类型：模糊 vs.精确) × 2 (反馈效价：积极 vs.消极) × 5 (脑区：额区 vs.额中区 vs.中央区 vs.中顶区 vs.顶区)的重复测量方差分析发现, 反馈类型具有主效应, $F(1, 34) = 4.98$, $p = 0.03$, $\eta_p^2 = 0.16$, 95% CI = [1.17, 4.45]; 在模糊反馈条件下出现自我欺骗的 P2 (2.81 μV)显著大于精确反馈条件(1.09 μV)。反馈效价没有主效应, $F(1, 34) = 1.64$, $p = 0.21$。脑区的主效应极其显著, $F(4, 34) = 58.74$, $p < 0.001$, $\eta_p^2 = 0.69$。事后多重比较可知, 在额区(4.28 μV)诱发的 P2 成分显著大于中央区(2.38 μV), $p < 0.001$, 95% CI = [0.51, 3.28]; 显著大于中顶区(0.43 μV), $p < 0.001$, 95% CI= [1.96, 5.73]和顶区(−1.01 μV), $p < 0.001$, 95% CI = [3.15, 7.43]; 但是与额中区(3.69 μV)无显著差异, $p = 0.12$。反馈类型和反馈效价不存在交互作用, $F(3, 34) = 0.69$, $p = 0.41$; 反馈类型和脑区不存在交互作用, $F(9, 34) = 1.55$, $p = 0.22$; 反馈效价和脑区不存在交互作用, $F(9, 34) = 0.15$, $p = 0.71$。反馈类型、反馈效价和脑区不存在交互作用, $F(19, 34) = 0.12$, $p = 0.74$ (见图 9)。

通过对 P2 成分进行 2 (反馈类型：模糊 vs.精确) × 2 (反馈效价：积极 vs.消极) × 3 (脑半球：左半球 vs.中央区 vs.右半球)的重复测量方差分析发现, 反馈类型的主效应边缘显著, $F(1, 34) = 3.31$, $p = 0.08$, $\eta_p^2 = 0.10$, 95% CI = [0.85, 3.93], 在模糊反馈条件下出现自我欺骗的 P2 (2.39 μV)显著大于精确反馈条件(1.08 μV)。反馈效价的主效应不显著, $F(1, 34) = 1.92$, $p = 0.18$, 脑半球的主效应不显著, $F(2, 34) = 1.99$, $p = 0.17$。反馈类型和反馈效价不存在交互作用, $F(3, 34) = 0.71$, $p = 0.41$。

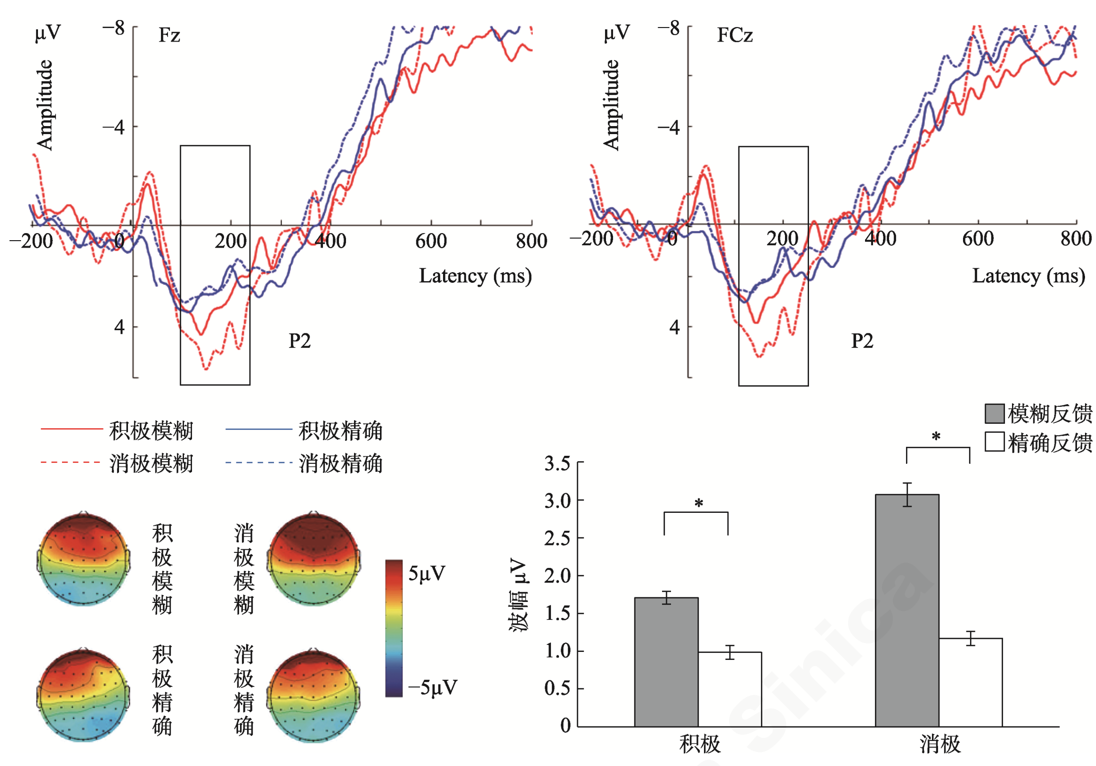


图 9　四个条件在 Fz 和 FCz 电极点的 P2 (时间窗：100~250 ms)的 ERP 总平均图、脑地形图和所有分析点的 P2 成分平均值柱状图, $*p < 0.05$;$**p < 0.01$;$***p < 0.001$

反馈类型和脑半球存在交互作用, $F(5, 34) = 6.31$, $p = 0.02$, $\eta_p^2 = 0.18$; 简单效应分析可知, 在右半球, 相比于精确反馈(0.46 μV), 模糊反馈(2.14 μV)在诱发 P2 成分上具有显著差异, $F(1, 34) = 4.63$, $p = 0.04$, 95% CI = [0.85, 3.43]; 在中央区, 相比于精确反馈(1.26 μV), 模糊反馈(2.71 μV)在诱发 P2 成分上具有显著差异, $F(1, 34) = 3.7$, $p = 0.06$, 95% CI = [1.06, 4.36]; 在左半球, 相比于精确反馈, 模糊反馈不具有显著, $F(1, 34) = 1.74$, $p = 0.19$。反馈效价和脑半球不存在交互作用, $F(5, 34) = 0.62$, $p = 0.43$。反馈类型、反馈效价和脑半球三者不存在交互作用, $F(19, 34) = 1.48$, $p = 0.23$。

(3) N400 成分

通过对 N400 成分进行 2 (反馈类型：模糊 vs. 精确) × 2(反馈效价：积极 vs.消极) × 5 (脑区：额区 vs.额中区 vs.中央区 vs.中顶区 vs.顶区)的重复测量方差分析发现, 反馈类型没有主效应, $F(1, 34) = 1.99$, $p = 0.17$。反馈效价没有主效应, $F(1, 34) = 0.02$, $p = 0.89$。脑区具有主效应, $F(4, 34) = 14.04$, $p = 0.001$, $\eta_p^2 = 0.33$。事后多重比较可知, 在额区(−4.66 μV)诱发的 N400 成分显著小于中央区(−3.87 μV), $p = 0.07$, 95% CI = [−3.12, −0.07]; 中顶区(−2.43 μV), $p = 0.04$, 95% CI = [−4.42, −0.06]; 和顶区(−1.43 μV), $p = 0.005$, 95% CI = [−5.76, −0.71]; 但是与额中区(−4.07 μV)无显著差异, $p = 1.00$。反馈类型和反馈效价不存在交互作用, $F(3, 34) = 0.48$, $p = 0.49$。反馈类型和脑区不存在交互作用, $F(9, 34) = 0.75$, $p = 0.39$。反馈效价和脑区不存在交互作用, $F(9, 34) = 0.43$, $p = 0.52$。

反馈类型、反馈效价和脑区三者有交互作用, $F(19, 34) = 2.86$, $p = 0.05$, $\eta_p^2 = 0.29$。通过简单简单效应的分析发现, 在消极反馈条件下, 相比于模糊反馈, 精确反馈在额区边缘显著的诱发 N400 成分, $F(9, 34) = 3.01$, $p = 0.09$, 95% CI = [−5.74, −1.64]; 在额中区边缘显著的诱发 N400 成分, $F(9, 34) = 2.93$, $p = 0.09$, 95% CI = [−4.97, −1.00]; 其他的条件下, 均为无显著差异, $ps > 0.05$ (见图 10)。

通过对 N400 成分进行 2 (反馈类型：模糊 vs. 精确) × 2 (反馈效价：积极 vs.消极) × 3 (脑半球：左半球 vs.中央区 vs.右半球)的重复测量方差分析发现, 反馈类型的主效应不显著, $F(1, 34) = 1.95$, $p = 0.17$。反馈效价的主效应也不显著, $F(1, 34) = 0.002$, $p = 0.96$。脑半球的主效应也不显著, $F(2, 34) = 0.54$, $p = 0.59$。反馈类型和反馈效价不存在交互作用, $F(3, 34) = 0.61$, $p = 0.44$。反馈类型和脑半球不存在交互作用, $F(5, 34) = 0.13$, $p = 0.88$。反馈效价和脑半球不存在交互作用, $F(5, 34) = 0.15$, $p = 0.23$。反馈类型、反馈效价和脑半球三者不存在交

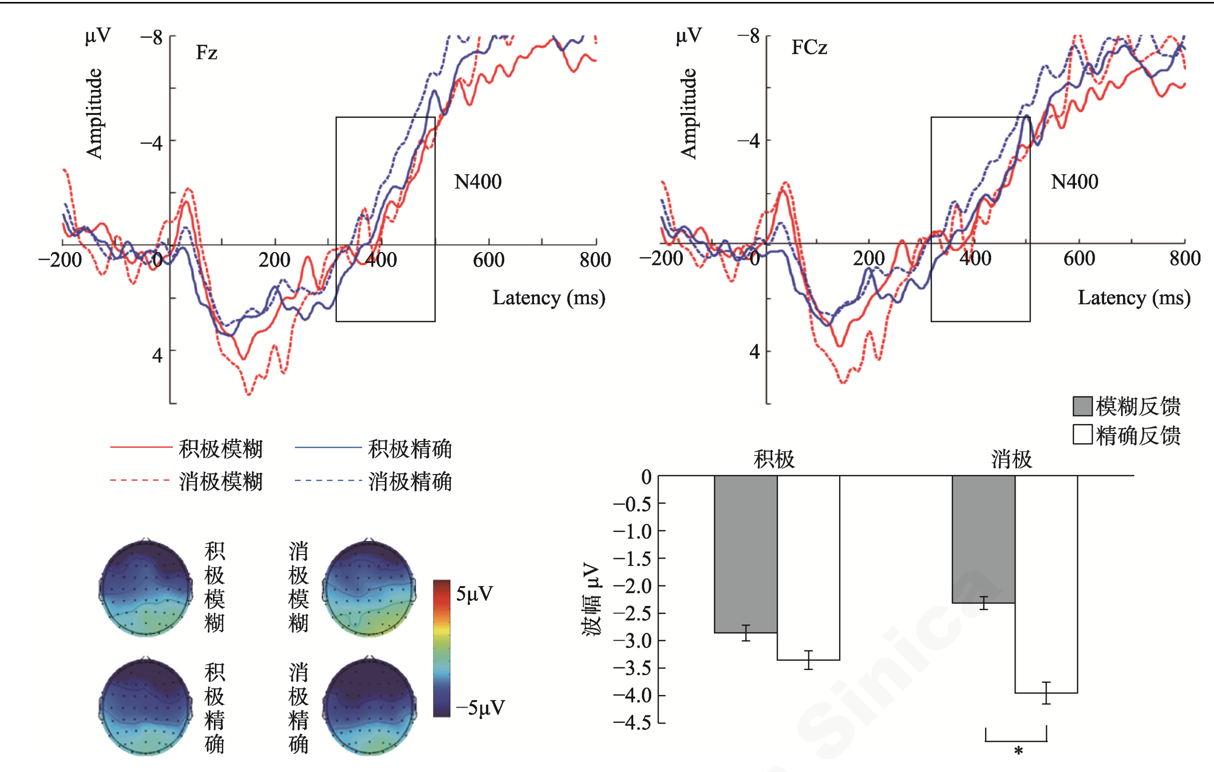


图 10 四个条件在 Fz 和 FCz 电极点的 N400 (时间窗：300~500 ms)的 ERP 总平均图、脑地形图和所有分析点的 N400 成分平均值柱状图, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

互作用, $F(19, 34) = 0.90$, $p = 0.41$。

4.3.5 行为结果与 ERPs 的相关

通过皮尔逊积差相关分析发现, 测试 2 第一阶段的预测分数与 N400 成分存在显著的负相关, $r = -0.41$, $p = 0.03$。这表明反馈诱发的 N400 成分越小, 被试在测试 2 第一阶段的预测分数越大(见图 11)。通过皮尔逊积差相关分析发现, 测试 2 第一阶段的预测分数与 N1 和 P2 成分没有显著的相关性, $ps > 0.05$。

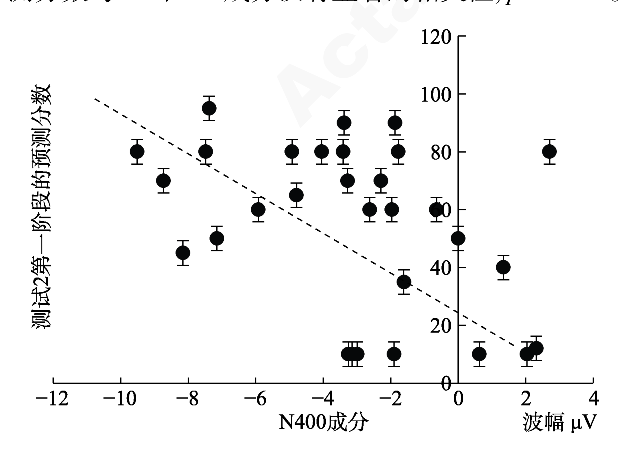


图 11 测试 2 第一阶段的预测分数与 N400 成分的相关分析

4.4 讨论

研究结果发现个体在模糊条件下产生更多的自我欺骗次数, 模糊反馈会让个体对自己的能力产生更多的解释, 积极反馈使个体对信息进行有利于自己的选择和评价。这些研究结果可能表明个体在任务中获得的积极模糊反馈的信息对自我欺骗产生具有更好的促进作用。

另外, 研究结果发现个体在积极模糊反馈条件可以在额区、额中区、中央区和中顶区诱发更小的 N1 成分, 这样的研究表明积极模糊反馈可以在个体的大脑中产生较大的作用以及自我欺骗在额区被显著的激活(Schapkin et al., 2000; Stahl et al., 2010)。模糊反馈诱发更多的 P2 成分, P2 成分对有关情绪的信息比较敏感, 这往往反映了大脑对情绪的整体评价。本研究结果也表明模糊反馈信息诱发自我欺骗的活跃区是在右半球以及中央区, 因为右半球是主要控制人们感性、情感、态度以及想象等功能。个体在模糊信息反馈的作用下在额区和额中区诱发更小的 N400 成分, 可能是个体在模糊反馈下需要选择一致的语义导致诱发 N400 成分较低, 相反在精确反馈下的个体选择不一致的语义导致诱发的 N400 成分较高(Näätänen et al., 2007)。另外通过皮尔逊积差相关分析可知, N400 与测试 2 第一阶段的预测分数具有显著的负相关, 这表明个体比较在意实验程序中的语句是否一致, 自我欺骗的产生可能不需要较大的语义冲突, 这表明自我欺骗是一种内在的状态(Wang et al., 2017)。

5　总讨论

5.1　模糊信息促进自我欺骗的产生

3 个实验都发现模糊反馈对自我欺骗的产生具有促进作用，这表明个体在模棱两可或主观性较强的情况下会积极保持正面的社会或自我形象(Dana et al., 2006; Kunda, 1990)。实验 1 的个体通过测试 1 的启动产生自我欺骗，在面对无论是积极模糊还是消极模糊信息时，个体在测试 2 第二阶段都预测了自己具有较高的分数(Haisley & Weber, 2010; Ren et al., 2018; Sloman et al., 2010)。在实验 2 产生自我欺骗的个体却对测试 2 具有较低的预测分数，这个原因可能说明金钱不会维持自我欺骗的稳定性；当个体真正进入自我欺骗后，即使对行为进行奖励也无法让其从自我欺骗中走出来(Chance et al., 2011)。本研究结合前两个实验的材料展开实验 3 的测试，行为结果也发现模糊反馈信息对产生自我欺骗个体的预测分数影响较大(Grossman & van der Weele, 2017)。该结果可能是因为模糊信息给个体提供了不同的解释方向，当评价特征涉及到模糊的属性时，模糊信息的反馈给个体提供想象的空间，使个体可能去夸大自己的能力(Gneezy et al., 2019; Liu et al., 2019; Haisley & Weber, 2010)。

实验3的ERP结果还发现模糊反馈在额区和额中区诱发较小的 N400 成分。本研究处于自我欺骗状态中的个体选择“大于”的语句没有诱发更大的 N400 成分，这说明个体没有产生较大的冲突，所以反过来证明个体处于自我欺骗的状态(Holroyd et al., 2006; Nieuwenhuis et al., 2005)。自我欺骗可以减少个体在实验过程中的认知负荷，主要是自我欺骗可以利用自身的特点，以及 N400 成分有助于个体在做语义任务时进行更快无冲突的判断(Neuber, 2016; Schwardmann & van der Weele, 2019)。总之，本研究发现：模糊信息的反馈促进了自我欺骗的产生。N400 成分在自我欺骗产生上的激活变化以及额区存在的作用。虽然有不少的学者对此进行了探索，但是该结果的稳定性和真实性还需要更多的学者不断去发展与证实。

5.2　积极信息推动效价在自我欺骗中的促进作用

根据以往的研究和本研究的结果发现自我欺骗的发生需要来自外界模糊信息的反应，因为未能明确地得出结论本身可能就是一种自我欺骗，这样可以允许人们在最大程度地说服自己这是他们能力所得到的结果(Sloman et al., 2010)。实验 1 和实验 3 发现积极效价在模糊反馈中对自我欺骗的效应要大于消极效价在模糊反馈中的效应。虽然实验 2 没有模糊信息的反馈，但是保留积极信息对自我欺骗的研究，实验 2 也发现积极信息对自我欺骗的作用较大的结果与以往的研究结论相一致(谭成慧 等, 2020)。

3 个实验都发现积极效价在自我欺骗中的作用，主要表现在模糊信息的基础上添加积极情感更有助于自我欺骗的产生。这可能是由于处于积极情绪下的个体分泌了较多的多巴胺，导致个体更容易产生认知上的偏差，对自我抱有一种积极乐观的信念，并通过自我欺骗去维持个人的理想形象(Foster & Frijters, 2014)。通过实验 3 的 ERP 结果可知，自我欺骗主要激活 P2 成分，这与个体比较在意积极效价的结果相一致(Rottenburger et al., 2019)。有研究表明 P2 成分是靶刺激基于相对简单的刺激特征进行定义，是与人们的情感和动机有关。这样的研究结果证明了行为实验操作的有效性，不仅从正面证实了自我欺骗是积极效价的作用，而且也从侧面证实了自我欺骗与积极情感的关系。实验 3 更是从内在神经机制层面证实了自我欺骗主要是在积极效价作用下发生的。总之，本研究的第二个研究发现是：在模糊信息的传递中若带有积极效价，个体产生自我欺骗的概率会更大。

5.3　前瞻范式下自我欺骗的背后机制可能指向了自我膨胀

3 个实验都发现个体在测试 1 中因有外在答案的提示在测试 2 的预测分数往往要高于没有外在答案提示的个体。这些结果表明个体在测试 1 所形成的“虚假自信”往往会在预测接下来的任务中表现出“虚假高分”。他们就会将这种“虚假好表现”延续到将来的任务中从而就形成一种“自我膨胀”(Dijkstra et al., 2012; von Hippel & Trivers, 2011)。3 个实验也发现答案提示组的个体都在测试 2 第二阶段提高预测分数，这可能表明个体在前期所形成的“自我膨胀”并没有在短暂的时间内消散，个体将这种“自我膨胀”所带来的状态影响接下来任务的完成(Bayne & Fernández, 2009)。这些结果可能表明了自我欺骗所带来的自我膨胀具有延长性。

个体得到无限膨胀以及失去真实的自我，这种“自我膨胀”的自我欺骗可能会导致个体产生错误的认知偏差和错误的处事观念等方面的变化(Lopez & Fuxjager, 2012; 鞠实儿 等, 2003)。虽然本研究对相关研究还没有进行探讨，但是这可以为

将来的研究工作提供新的探索方向。为自我欺骗是否带来消极心理作用的背后机制及干预的探究提供思路，以帮助个体更好地监控自我欺骗行为和规避个体的损失，最后减轻自我欺骗对社会产生的危害作用。

5.4 研究不足与展望

本研究采用“前瞻范式”的方法对自我欺骗进行探索，发现了积极模糊信息在自我欺骗中的促进作用，这为自我欺骗的研究增加证据。但是在本研究中也存在一些问题。首先，针对实验 2 的结果没有效果的原因可能是因为，本研究在精确反馈的条件中主要采用是较少金钱，金钱的启动也许无法更好的激发个体的兴趣；希望在未来的研究中提高金钱的额度。若实验结果发现即便具有较大额度的金钱反馈也不会使个体产生更多的自我欺骗，这就可以从侧面证实影响自我欺骗最大的因素来自模糊反馈。其次，本研究主要探索自我欺骗的存在现状和影响自我欺骗产生的因素。虽然得出一定有价值的结论，但是本研究还有很多因素没有考察到，比如采用更加科学的 fMRI 去探索自我欺骗被激活的内在神经机制。最后，本研究主要选取的是实验对象比较单一的在校大学生，这个结果没有办法扩展到其他年龄阶段。所以在未来的研究中，我们希望可以将实验样本进行扩大化，去探索青少年、老年阶段的个体是否会产生自我欺骗的相似结论。

参 考 文 献

Bandura, A. (2011). Self-deception: A paradox revisited. *Behavioral Brain Sciences, 34*(1), 16–17.

Bayne, T., & Fernández, J. (2009). Delusion and self-deception: Mapping the terrain. In T. Bayne & J. Fernandez (Eds.), *Delusion and self-deception: Affective and motivational influences on belief formation* (pp. 1–21). Psychology Press.

Chance, Z., Gino, F., Norton, M. I., & Ariely, D. (2015). The slow decay and quick revival of self-deception. *Frontiers in Psychology, 6*, 1075.

Chance, Z., & Norton, M. I. (2015). The what and why of self-deception. *Current Opinion in Psychology, 6*, 104–107.

Chance, Z., Norton, M. I., Gino, F., & Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. *Proceedings of the National Academy of Sciences*, *108*(Suppl. 3), 15655–15659.

Chen, T. B., & Chai, L. T. (2010). Attitude towards the environment and green products: Consumers' perspective. *Management Science and Engineering*, *4*(2), 27–39.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, *100*(2), 193–201.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Dijkstra, P. D., Schaafsma, S. M., Hofmann, H. A., & Groothuis, T. G. (2012). 'Winner effect' without winning: Unresolved social conflicts increase the probability of winning a subsequent contest in a cichlid fish. *Physiology & Behavior*, *105*(2), 489–492.

Duran, K. A., O'Halloran, H., Soder, H., Yasin, S., Kramer, R., Rosen, S., Brenya, J., Chavarria, K., Savitska, L., & Keenan, J. P. (2020). The medial prefrontal cortex: A potential link between self-deception and affect. *International Journal of Neuroscience*, *131*(7), 701–707.

Egerton, A., Rees, E., Bose, S. K., Lappin, J. M., Stokes, P., & Turkheimer, F. E., Reeves, S. J. (2010). Truth, lies or self-deception? striatal d (2/3) receptor availability predicts individual differences in social conformity. *NeuroImage*, *53*(2), 777–781.

Fan, W., Yang, B., Liu, J., & Fu, X. (2017). Self-deception: For adjusting individual psychological states. *Advances in Psychological Science*, *25*(8), 1349–1359.

[范伟, 杨博, 刘娟, 傅小兰. (2017). 自我欺骗 为了调节个体心理状态. *心理科学进展, 25*(8), 1349–1359.]

Farrow, T. F. D., Burgess, J., Wilkinson, I. D., & Hunter, M. D. (2015). Neural correlates of self-deception and impression-management. *Neuropsychologia, 67*, 159–174.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 202–224.

Foster, G., & Frijters, P. (2014). The formation of expectations: Competing theories and new evidence. *Journal of Behavioral and Experimental Economics*, *53*, 66–81.

Galeotti, A. E. (2016). Straight and twisted self-deception. *Phenomenology and Mind*, *11*, 90–99.

Gao, J. (2021). Self–deception and pragmatic encroachment: A dilemma for epistemic rationality. *Ratio*, *34*(1), 20–32.

Gneezy, U., Saccardo, S., & van Veldhuizen, R. (2019). Bribery: Behavioral drivers of distorted decisions. *Journal of the European Economic Association*, *17*(3), 917–946.

Grossman, Z., & van der Weele, J. J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, *15*(1), 173–217.

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality & Social Psychology, 37*(2), 147–169.

Haisley, E. C., & Weber, R. A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games & Economic Behavior, 68*(2), 614–625.

Holroyd, C. B., Hajcak, G., & Larsen, J. T. (2006). The good, the bad and the neutral: Electrophysiological responses to feedback stimuli. *Brain Research*, *1105*(1), 93–101.

Jian, Z., Zhang, W., Tian, L., Fan, W., & Zhong, Y. (2019). Self-deception reduces cognitive load: The role of involuntary conscious memory impairment. *Frontiers in Psychology*, *10*.

Ju, S. E., Zhao, Y., & Fu, X. (2003). Cognitive mechanism of self- deception. *Journal of Sun Yat-sen University (Social Science Edition)*, *43*(5), 19–26.

[鞠实儿, 赵艺, 傅小兰. (2003). 自我欺骗的认知机制. *中山大学学报(社会科学版), 43*(5), 19–26.]

Kunda, Z. (1990). The case for motivated reasoning. *Psychological*

Bulletin, 108(3), 480–498.
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.
Lai, V. T., & Huettig, F. (2016). When prediction is fulfilled: Insight from emotion processing. *Neuropsychologia*, *85*, 110–117.
Lauria, F., & Preissmann, D. (2018). What does emotion teach us about self-deception? Affective neuroscience in support of non-intentionalism. In *Les ateliers de l'éthique/The Ethics Forum* (Vol. 13, No. 2, pp. 70–94*)*. Centre de recherche en éthique de l'Université de Montréal.
Liu, J., Zhang, W., Zhan, Y., Song, L., Guan, P., Kang, D. Jian, J., Cai, R., & Li, M. (2019). The effect of negative feedback on positive beliefs in self-deception. *Frontiers in Psychology, 10*, 702.
Liu, Y., & Shu, H. (2003). ERP and language studies. *Advances in Psychological Science*, *11*(3), 296–302.
[刘燕妮, 舒华. (2003). ERP 与语言研究. *心理科学进展*, *11*(3), 296–302.]
Lopez, J. K., & Fuxjager, M. J. (2012). Self-deception's adaptive value: Effects of positive thinking and the winner effect. *Consciousness and Cognition*, *21*(1), 315–324.
Mele, A. R. (1997a). Real self-deception. *Behavioral & Brain Sciences, 20*, 91–136.
Mele, A. R. (1997b). Understanding and explaining real self-deception. *Behavioral Brain Sciences, 20*(1), 127–134.
Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544–2590.
Neuber, S. (2016). Self-awareness and self-deception: A Sartrean perspective. *Continental Philosophy Review*, *49*(4), 485–507.
Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532.
Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, *19*(2), 209–213.
Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: Identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience, 6*, 278.
Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality & Social Psychology, 46*(2), 237–248.
Ren, M., Zhong, B., Fan, W., Dai, H., Yang, B., Zhang, W., Yin, Z., Liu, J., Li, J., & Zhan, Y. (2018). The influence of self-control and social status on self-deception. *Frontier in Psychology, 9*, 1256.
Rottenburger, J. R., Carter, C. R., & Kaufmann, L. (2019). It's alright, it's just a bluff: Why do corporate codes reduce lying, but not bluffing?. *Journal of Purchasing and Supply Management*, *25*(1), 30–39.
Schapkin, S. A., Gusev, A. N., & Kuhl, J. (2000). Categorization of unilaterally presented emotional words: An ERP analysis. *Acta Neurobiologiae Experimentalis*, *60*(1), 17–28.
Schwardmann, P., & van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, *3*(10), 1055–1061.
Sharma, M., Modi, S., Khushu, S., & Mandal, M. K. (2010). Neural activation pattern in self-deceivers. *Psychological Studies*, *55*(1), 71–76.
Sheridan, Z., Boman, P., Mergler, A., & Furlong, M. J. (2015). Examining well-being, anxiety, and self-deception in university students. *Cogent Psychology*, *2*(1), 993850.
Sloman, S. A., Fernbach, P. M., & Hagmayer, Y. (2010). Self-deception requires vagueness. *Cognition*, *115*(2), 268–281.
Smith, M. K., Trivers, R., & von Hippel, W. (2017). Self-deception facilitates interpersonal persuasion. *Journal of Economic Psychology*, *63*, 93–101.
Stahl, J., Wiese, H., & Schweinberger, S. R. (2010). Learning task affects ERP-correlates of the own-race bias, but not recognition memory performance. *Neuropsychologia*, *48*(7), 2027–2040.
Surbey, M. K. (2011). Adaptive significance of low levels of self-deception and cooperation in depression. *Evolution and Human Behavior, 32*(1), 29–40.
Tan, C., Zhu, C., & Song, H. (2020). The effect of emotional valency and emotional arousal on self–deception. *Journal of Neijiang Normal University*, *35*(10), 1–6.
[谭成慧, 朱传林, 宋欢. (2020). 情绪效价和情绪唤醒度对自我欺骗的影响. *内江师范学院学报*, *35*(10), 1–6.]
Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193–210.
Uziel, L., & Cohen, B. (2020). Self-deception and discrepancies in self-evaluation. *Journal of Research in Personality*, *88*, 1–5.
von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral & Brain Science, 34*(1), 1-16; discussion 16–56.
Wang, Y., Kuhlman, D. M., Roberts, K., Yuan, B., Zhang, Z., Zhang, W., & Simons, R. F. (2017). Social value orientation modulates the FRN and P300 in the Chicken Game. *Biological Psychology*, *127*, 89–98.
Zhong, L., & Mo, L. (2019). A commentary on empirical study of self-deception. *Journal Psychological Science*, *42*(3), 709–714.
[钟罗金, 莫雷. (2019). 自我欺骗实证研究述评. *心理科学*, *42*(3), 709–714.]
Zhong, L., & Ru, T. (2021). A Cognitive processing model of self-deception. *Journal of South China Normal University (Social Science Edition)*, (3), 81–89.
[钟罗金, 汝涛涛. (2021). 自我欺骗的认知加工: 模型及其构建. *华南师范大学学报 (社会科学版)*, (3), 81–89.]
Zhong, L., Ru, T., Fan, M., & Mo, L. (2019). The effect of cognitive vagueness and motivation on conscious and unconscious self-deception. *Acta Psychologica Sinica*, *51*(12), 1330–1340.
[钟罗金, 汝涛涛, 范梦, 莫雷. (2019). 认知模糊程度和动机强度对有意识和无意识自我欺骗的影响. *心理学报*, *51*(12), 1330–1340.]
Zhu, J., & Jiang, J. (2014). Self-deception: A new topic in contemporary psychology. *Science of Social Psychology*, *2*, 53–58.
[朱晶, 蒋京川. (2014). 自欺: 当代心理学研究新主题. *社会心理科学, 2*, 53–58.]
Zhu, Y., Xi, S., & Wu, Y. (2016). Self-deception in autobiographical narrative: Individuals' self-identity strategy. *Advances in Psychological Science*, *24*(12), 1917–1925.
[朱艳丽, 席思思, 吴艳红. (2016). 自传体叙事中的自我欺骗: 个体的自我认同策略. *心理科学进展*, *24*(12), 1917–1925.]

The impact of feedback on self-deception: Evidence from ERP

FAN Wei[1,2], REN Mengmeng[1,2], ZHANG Wenjie[2,3], ZHONG Yiping[1,2]

([1] *Department of Psychology, Hunan Normal University, Changsha* 410081, *China*)

([2] *Cognition and Human Behavior Key Laboratory of Hunan Province, Changsha* 410081, *China*)

([3] *Department of Preschool Education, Hunan Normal University, Changsha* 410081, *China*)

Abstract

Self-deception is a motivation to see the world in a way that benefits oneself. Previously, researchers have explored self-deception from a single research perspective. The focus on singular variables have made it impossible to understand an individual's processing of self-deception from a holistic perspective. Researchers of self-deception not only need to seek stable research results but also need to continuously adopt a more comprehensive research perspective to explore self-deception.

This study mainly explored the influence of feedback types and feedback valence on self-deception. Experiment 1 used the forward-looking paradigm to explore the influence of fuzzy feedback with differing valences on self-deception. The scores (measures of self-deception) in the first stage of test 2 in the answer prompt group and the no-answer prompt group were significantly different. The prediction scores in the second stage of test 2 were also significantly different, and compared with the participants in the no-answer prompt group, the participants in the answer prompt group increased their prediction scores in the second stage of test 2 under conditions involving positive fuzzy feedback. Experiment 2 explored the impact of precise feedback with differing valences on self-deception. The scores for the two groups of participants in the first stage of test 2 were significantly different. Compared with the no-answer prompt group, the answer prompt group received accurate feedback; however, these conditions did not produce higher prediction scores in the second stage of test 2. Experiment 3 used ERP technology to explore the internal mechanisms related to different types of feedback and the feedback valence that had differential effects on self-deception. The participants engaged in more self-deception in conditions with fuzzy feedback; self-deception in conditions with fuzzy feedback induced a larger P2 component and induced smaller N1 and N400 components. Finally, it was found that the frontal area had the greatest effects on self-deception.

In short, the results of this research showed that positive feedback promoted the generation of individuals' self-deception, while negative feedback weakened the generation of individuals' self-deception; positive fuzzy feedback further promoted the generation of individuals' self-deception. The results of this research also indicated that the mechanism underlying self-deception may be self-inflation.

Key words self-deception, feedback type, feedback valence, forward-looking paradigm, event-related potential